

Bradfordizing als Re-Ranking-Ansatz in Literaturinformationssystemen

Mayr, Philipp

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Mayr, P. (2011). Bradfordizing als Re-Ranking-Ansatz in Literaturinformationssystemen. *Information - Wissenschaft und Praxis*, 62(1), 1-15. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-220364>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

Bradfordizing als Re-Ranking-Ansatz in Literaturinformationssystemen

Philipp Mayr¹

Abstract

In diesem Artikel wird ein Re-Ranking-Ansatz für Suchsysteme vorgestellt, der die Recherche nach wissenschaftlicher Literatur messbar verbessern kann. Das nicht-textorientierte Rankingverfahren Bradfordizing wird eingeführt und anschließend im empirischen Teil des Artikels bzgl. der Effektivität für typische fachbezogene Recherche-Topics evaluiert. Dem Bradford Law of Scattering (BLS), auf dem Bradfordizing basiert, liegt zugrunde, dass sich die Literatur zu einem beliebigen Fachgebiet bzw. -thema in Zonen unterschiedlicher Dokumentenkonzentration verteilt. Dem Kernbereich mit hoher Konzentration der Literatur folgen Bereiche mit mittlerer und geringer Konzentration. Bradfordizing sortiert bzw. rankt eine Dokumentmenge damit nach den sogenannten Kernzeitschriften. Der Retrievaltest mit 164 intellektuell bewerteten Fragestellungen in Fachdatenbanken aus den Bereichen Sozial- und Politikwissenschaften, Wirtschaftswissenschaften, Psychologie und Medizin zeigt, dass die Dokumente der Kernzeitschriften signifikant häufiger relevant bewertet werden als Dokumente der zweiten Dokumentzone bzw. den Peripherie-Zeitschriften. Die Implementierung von Bradfordizing und weiteren Re-Rankingverfahren liefert unmittelbare Mehrwerte für den Nutzer.

This article proposes a re-ranking approach for search systems which intends to improve the retrieval for scientific literature on a significant level. The non text-oriented ranking approach Bradfordizing will be introduced and then the effectiveness for typical subject specific retrieval topics will be evaluated. The underlying Bradford Law of Scattering (BLS) states that the literature on any subject or domain is distributed in different zones with typical concentrations in each zone. The core zone or core with a high concentration of literature is followed by zones with decreasing and low concentrations of papers. Bradfordizing sorts or ranks a document space by the so called core journals. The retrieval test with 164 intellectually assessed topics from different abstracting and indexing databases in the social and political sciences, economics, psychology and medical science shows that documents in the core journals are significantly more often relevant than documents from zone 2 journals and peripheral journals. The paper shows that the implementation of Bradfordizing and other re-ranking approaches generates an immediate value-added for users.

1 Einleitung

Nutzer sehen sich je nach Art und Umfang der Literaturrecherche immer häufiger in der Situation, auf unterschiedliche Suchsysteme zugreifen zu müssen, um ihren Informationsbedarf zu befriedigen. Inzwischen führt die alleinige Nutzung z. B. eines Online-Katalogs (OPAC) der lokalen Bibliothek genauso wenig zu vollständigen Ergebnissen wie die Beschränkung bei der Suche auf eine einzelne Fachdatenbank. Vor allem wissenschaftliche Nutzer erwarten aus diesem Grund heute möglichst die Bündelung und Konzentration fachlich relevanter und qualitativer Information an einer Stelle.

¹ Email: philipp.mayr@gesis.org

Des Weiteren fordern Nutzer für datenbankübergreifende Literaturrecherchen mit großen Dokumentmengen einen möglichst hohen Anteil an relevanten und qualitativ hochwertigen Dokumenten in den vorderen Trefferergebnissen. Insbesondere die Reihenfolge und Struktur der gelisteten Ergebnisse (Ranking) spielt, neben dem direkten Volltextzugriff auf die Dokumente, inzwischen eine entscheidende Rolle beim Design von Suchsystemen. Abgegrenzt wird Ranking oder Relevance Ranking von sog. Sortierungen z. B. nach dem Erscheinungsjahr der Publikation, obwohl hier die Grenze zu „nach inhaltlicher Relevanz“ gerankten Listen konzeptuell nicht sauber zu ziehen ist. Das Ranking von Dokumenten führt letztlich dazu, dass sich die Benutzer solcher Suchsysteme fokussiert mit den oberen Treffermengen (top-gerankt) eines Suchergebnisses beschäftigen. Der mittlere und untere Bereich eines Suchergebnisses wird häufig nicht mehr in Betracht bezogen. Bei großen Treffermengen ist eine vollständige Rezeption eines Suchergebnisses ohnehin nicht mehr möglich. Häufig ist die Konsequenz, dass Nutzer aus Bequemlichkeitsgründen ausschließlich mit kommerziellen Internetsuchmaschinen wie Google oder Yahoo recherchieren und die fachwissenschaftlich erschlossenen Literaturdatenbanken nicht mehr konsultiert werden.

Auf Grund der Vielzahl an relevanten und verfügbaren Informationsquellen ist es daher notwendig, Kernbereiche in den Suchräumen zu identifizieren und diese anschließend dem Nutzer hervorgehoben zu präsentieren. Die Kernbereichsbildung kann zum einen über unscharfe textstatistische Ranking-Verfahren (z. B. term frequency - inverse document frequency (tf-idf)) erfolgen, zum anderen aber auf Grundlage der den Suchräumen inhärenten Strukturen.

Das in diesem Artikel evaluierte Verfahren Bradfordizing setzt auf einer typischen bibliometrischen Regelmäßigkeit thematischer Dokumentenräume auf und nutzt diese Strukturmerkmale zum Ranking von Dokumenten (siehe dazu Mayr, 2010).

2 Motivation

Neben der Abgrenzung der Suchräume und Dokumenttypen erhält die Zugänglichkeit und Relevanz der Dokumente eine entscheidende Bedeutung für den Benutzer. Die neueren technologischen Entwicklungen des Web Information Retrieval (IR), wie sie die kommerziellen Internetsuchmaschinen implementieren, werten insbesondere frei zugängliche Dokumente mit ihrer gesamten Text- und Linkinformation automatisch aus. Diese Verfahren sind vor allem deshalb erfolgreich, weil sie Ergebnislisten nach Relevanz gerankt darstellen, einfach und schnell zu recherchieren sind und direkt auf die Informationsquellen verweisen. Die qualitativen Verfahren der traditionellen Fachinformationsanbieter (z. B. Web of Science) hingegen zeigen genau bei diesen Punkten (Ranking, Einfachheit und Volltextzugriff) Schwächen, überzeugen aber vor allem durch ihre Stringenz, in diesem Fall die selektive Aufnahme von qualitätsgeprüften Dokumenten in das System und die inhaltliche Erschließung der Dokumente. Aufgrund der Beschaffenheit der Dokumente in klassischen Informationssystemen wie z. B. bibliographischen Fachdatenbanken greifen die Ranking-Verfahren des Web IR genauso wenig wie die herkömmlichen textstatistischen Information Retrieval-Verfahren. Alternative, nicht-textorientierte Ansätze, wie z. B. der Einsatz von Autorenzentralitätsmaßen beim Ranking (Mutschke, 2004), aber auch das in diesem Artikel vorgestellte Verfahren, zeigen, dass die spezifische Struktur der Literaturnachweise und die Qualität der Daten (i. d. R. Metadaten) für alternative Mehrwertdienste gewinnbringend eingesetzt werden können.

Das Mehrwertverfahren Bradfordizing zielt auf die Neu- bzw. Reorganisation eines vorliegenden Trefferergebnisses (sog. Re-Ranking) und setzt damit konzeptuell nach der Suche bzw. Suchraumerweiterung z.

B. durch Heterogenitätsbehandlung (vgl. Krause et al, 2003; Krause, 2004; Mayr & Petras, 2008) an. In diesem Artikel soll überprüft werden, ob sich automatisch generierte Kernbereiche der Forschungsliteratur für den Nutzer zur Informationsrecherche eignen. Es wird angenommen, dass die Kernzonen/Kernzeitschriften nach Bradfordizing (hohe Artikelhäufigkeit zu einem Thema) positive Auswirkungen für den Nutzer zeigen (Hood & Wilson, 2001).

Motivation dieses Artikels ist es, empirisch zu untersuchen, ob das hier vorgestellte alternative Re-Rankingverfahren Bradfordizing auf Basis der Bradfordschen Gesetzmäßigkeit im Anwendungsbereich bibliographischer Datenbanken zum einen operabel ist und zum anderen voraussichtlich gewinnbringend in Informationssystemen eingesetzt und dem Nutzer angeboten werden kann.

3 Bradfordsche Gesetzmäßigkeit

Der folgende Abschnitt trägt die Grundlagen von Bradford's Law of Scattering (BLS) zusammen.

Der Begriff BLS stammt ursprünglich von Vickery (1948), der sich sehr früh mit den Grundlagen der Bradfordschen Gesetzmäßigkeit beschäftigt hat. Die Beobachtung des Bibliothekars Samuel Bradford aus den frühen 1930er Jahren kann als eine der Wurzeln der Informetrie (engl. Informetrics) bezeichnet werden. Die Informetrie ist wiederum eine Spezialdisziplin der Informationswissenschaft, die sich mit der mathematischen Beschreibung und Modellierung der Regularitäten bzw. sogenannten „Gesetzmäßigkeiten“ der beobachtbaren Objekte im Bibliotheks- und Informationswesen beschäftigt (vgl. Nacke, 1979). Zu diesen Objekten zählen z. B. Autoren, Publikationen (Zeitschriften, Monographien usw.), Referenzen, Zitationszahlen usw. Wichtig bei der Beurteilung der informetrischen Gesetzmäßigkeiten ist, dass diese keine Gesetzmäßigkeiten in Sinne von Naturgesetzen wie z. B. der Fallgesetze oder des Gravitationsgesetzes darstellen, sondern empirische Gesetze (Wilkinson, 1972). Diese können zwar mathematisch über Formeln beschrieben werden, Abweichungen und Unschärfe bei der Formulierung und Beobachtung der informetrischen Zusammenhänge sind aber die Norm. Häufig können die informetrischen Modelle die Realität sehr gut abbilden, weshalb die Bezeichnung Gesetzmäßigkeit naheliegt.

Dem BLS liegt zugrunde, dass sich die Literatur zu einem beliebigen Fachgebiet bzw. -thema, z. B. in einer Bibliographie, in Bereiche mit unterschiedlichen Dokumentenkonzentrationen unterteilen lässt. So besteht zwischen den Zeitschriften eines Forschungsthemas und den Artikeln in diesen Zeitschriften eine quantifizierbare Relation, die Bradford als erstes beschrieben hat. Dem Kernbereich mit hoher Konzentration der Literatur folgen Bereiche mit zunächst mittlerer und geringer Konzentration, die jeweils die gleiche Menge an Zeitschriftenartikeln beinhalten wie der Kernbereich. Bradford beschreibt diese Beobachtung mit:

“The whole range of periodicals thus acts as a family of successive generations of diminishing kinship, each generation being greater in number than the preceding, and each constituent of a generation inversely according to its degree of remoteness.” (Bradford, 1934)

Bradford hat das nach ihm benannte informetrische Gesetz wie folgt formuliert:

“... if scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as $1 : n : n^2 \dots$ ” (Bradford, 1934, S. 116)

Bradford selber hat in seinem Buch „Documentation“ ein Kapitel „The Documentary Chaos“ geschrieben und in diesem sein Verteilungsgesetz als Argument dafür genommen, Bibliographien künftig nach Zeitschriften auszuwerten und nicht nur nach rein fachlichen Gesichtspunkten aufzubauen. Bradford beschreibt damit sehr früh die Verfahren, mit denen noch heute Fachdatenbanken und Bibliographien aufgebaut werden. Bereits 1934 formuliert Bradford:

„It follows that the only way to glean all the articles on these subjects would be to scrutinise continually several thousands of journals, the bulk of which would yield only occasional references or none at all. And because this work is altogether impracticable, a large portion of the total number of references on a given subject is regularly missed by the abstracting journals. ... In methods of abstracting and indexing, a radical change is needed. Periodical literature must be abstracted by source, and not by subject, as hitherto.“ (Bradford, 1934, S. 179-180)

Die Besonderheit der Bradfordschen Beobachtung liegt vor allem in der Betonung der Rangfolge bzw. dem Ranking der Zeitschriften nach der Anzahl der Artikel zu dem untersuchten Thema. Das gleiche Verfahren wird bei den verwandten Gesetzmäßigkeiten Lotka's Law für Publikationen eines Autors sowie Worthäufigkeiten in Texten in Zipf's Law angewendet. BLS weist der produktivsten Zeitschrift den Rank 1 zu, der zweitproduktivsten Zeitschrift den Rank 2 usw. Auf dem letzten Rank befinden sich die Zeitschriften, die die geringste Anzahl an Artikeln zu dem Thema produzieren, das heißt, i. d. R. mindestens einen Artikel beinhalten.

Egghe und Rousseau (1990) nennen diese spezifischen Verteilungen „rank-order distributions“. Es werden in der Literatur aber auch Begriffe wie „long tail distributions“, „extremely skewed“, „law of the vital few“ oder auch „schiefe Verteilung“ verwendet. Power laws bzw. Potenzgesetze folgen im Übrigen der gleichen Verteilung (vgl. Newman, 2005).

Im Prinzip stellt das BLS lediglich eine Präzisierung der 80 : 20-Regel für den Bereich der Zeitschriftenliteratur dar. 80 % der Nachfrage an Literatur können mit etwa 20 % des Bestandes in einer Bibliothek abgedeckt werden (vgl. Umstätter, 2005; Wagner-Döbler, 1997). Anders formuliert: 80 % der Zeitschriftenartikel zu einem Thema finden sich in 20 % der Zeitschriften, die zu diesem Thema publizieren. BLS hat damit Parallelen zu Paretos Beobachtungen auf dem Gebiet der Einkommensverteilungen.

Folgendes Beispiel soll zur Illustration der Bradfordschen Gesetzmäßigkeit dienen. Angenommen, es werden 3.000 Artikel zu einem bestimmten Thema in insgesamt 800 Zeitschriften gefunden. Dann ergibt sich in der oben beschriebenen Reihung nach Bradford bei drei gleich großen Gruppen von jeweils 1.000 Artikeln ein Verhältnis von 42 : 158 : 600 Zeitschriften. Näherungsweise $1 : 3,8 : 3,8^2$

Der Core oder Kern der Zeitschriften besteht demnach aus insgesamt 42 Zeitschriften. Diese 42 Zeitschriften beinhalten die ersten 1.000 Artikel. In der zweiten Zone (Zone 2) befinden sich 158 Zeitschriften, die wiederum 1.000 Artikel beinhalten. Die 600 Zeitschriften aus der dritten Zone (Zone 3) gehören zu den Peripheriezeitschriften, die nur selten, aber wenn dann mindestens einen thematisch relevanten Artikel beinhalten. Die Konzentration der Zeitschriftenartikel in den Zeitschriften nimmt somit pro Zone konstant ab. Charakteristisch und allen Bradford-Verteilungen gemeinsam ist die auftretende Streuung (Ungleichverteilung) bzgl. der Artikelproduktivität einzelner Zeitschriften zwischen den Zonen, die sich annähernd wie ein Power law verhält.

Die Einteilung der beobachteten Objekte (Zeitschriften) in drei Zonen bei Bradford ist eine pragmatische Entscheidung, die keinen Zusammenhang mit der eigentlichen Gesetzmäßigkeit hat. Obwohl Bradford nur drei

Zonen bzw. Gruppen zugrundelegt, sind aufgrund der „Selbstähnlichkeit“ (self-similarity) dieser Verteilungen auch mehr Zonen möglich. Egghe und Rousseau (1990) zeigen dies an einem Beispiel.

Die umfangreiche Literatur zu BLS (siehe dazu Mayr, 2010) lässt sich in unterschiedliche Gruppen einteilen: a) die Bestätigung der Bradford-Gesetzmäßigkeit in unterschiedlichen Fachgebieten und Disziplinen, b) die Diskussion über die mathematisch geeignete und korrekte Formulierung der Gesetzmäßigkeit sowie die Beziehung von BLS zu anderen informatrischen Gesetzmäßigkeiten und c) die Diskussion der theoretischen und praktischen Implikationen des Gesetzes.

Bisher sind in der Literatur unterschiedliche Einsatzmöglichkeiten des BLS beschrieben, von denen hier nur einige genannt werden sollen. BLS dient dabei:

- zum Aufbau und Management sowie zur Kontrolle von Kollektionen,
- zur Auswahl von Zeitschriften für eine Fachdatenbank (z. B. bei Thomson Scientific; Garfield 1990;1996),
- zur Messung der Abdeckung von Bibliographien und
- zur Rechercheunterstützung (vor allem Bates, 2002, White, 1981).

Insbesondere die Einsatzmöglichkeiten des BLS für die Informationssuche bei Bates (Bates, 1989, Bates, 2002) sind relevant für diesen Artikel: "... the key point is that the distribution tells us that information is neither randomly scattered, nor handily concentrated in a single location. Instead, information scatters in a characteristic pattern, a pattern that should have obvious implications for how that information can most successfully and efficiently be sought." (Bates, 2002)

Es lassen sich unterschiedliche Szenarien denken, in denen ein Nutzer die Suche nur auf die Kernjournale eines Fachgebiets einschränken will (z. B. Überblicksrecherche). Im Sinne der Rechercheunterstützung kann davon ausgegangen werden, dass das Verteilungsmuster des BLS (Kernzone mit hoher Artikelhäufigkeit) für den Benutzer (vermutlich insbesondere ungeübte Nutzer und Anfänger) positive Auswirkungen beim Information Retrieval haben kann.

4 Bradfordizing

Das Verfahren Bradfordizing wurde erstmals in White (1981) zur Umorganisation von Suchergebnissen in Onlinedatenbanken vorgeschlagen. Bradfordizing sortiert bzw. rankt eine beliebige Dokumentmenge nach Kernzeitschriften. Die Zeitschriften zu einer Suchanfrage werden gemäß ihrer Häufigkeit (Anzahl der Zeitschriftenaufsätze) gelistet. Bradfordizing kann daher als ein nicht-textorientiertes Rankingverfahren bezeichnet werden. Aufsätze der Zeitschriften aus Zone 1 werden beim Bradfordizing vor die Zeitschriftenartikel geordnet, deren Zeitschriften nur durchschnittlich (Zone 2) oder selten (Zone 3) gelistet werden.

Bradfordizing ist ein Sortiervverfahren, das auf unterschiedliche Dokumenttypen in wissenschaftlichen Fachdatenbanken angewendet werden kann. White beschreibt das Verfahren folgendermaßen: "... That is sorting hits (1) by the journal in which they appear, and then sorting these journals not alphabetically by title but (2) numerically, high to low, by number of hits each journal contains. In effect, this two-step sorting ranks the search output in the classic Bradford manner, so that the most productive, in terms of its yield of hits, is placed first; the second-most productive journal is second; and so on, down through the last rank of journals yielding only one hit apiece." (S. 47).

Bradfordizing ist insofern ein innovatives Re-Ranking-Verfahren, als es für eine beliebige Suchanfrage über eine nachgeordnete Umorganisation des Suchergebnisses zuverlässig (hohe Robustheit dieser Häufigkeitsverteilung) die zentralen Publikationen für eine übergreifende Suche in der Ergebnisliste nach vorne sortiert, ohne die beschreibenden Textbestandteile der Dokumente zu analysieren. Bradfordizing „... erleichtert nicht nur das Auffinden von Aufsätzen und hilft Anfängern rasch zu erkennen, welche Zeitschriften für sie besonders wichtig sind, es erhöht auch die Wahrscheinlichkeit, gerade in diesen Quellen noch weitere wichtige Publikationen zu finden ...“ (Umstätter, 2005)

Der Ablauf einer Bradfordizing-Analyse lässt sich folgendermaßen zusammenfassen:

1. Identifikation von allen Dokumenten (Artikeln), die in einem Forschungsgebiet/Subject publiziert wurden, durch retrospektive Suche in einem ausreichend großen Zeitraum.
2. Auflistung der Quellen (Zeitschriften), in denen die Artikel publiziert wurden, in der Reihenfolge der Häufigkeit der Quelle. Das Ergebnis ist eine gerankte Liste der Zeitschriften (sog. „bradfordized list“).
3. Einteilung der Zeitschriftenliste in Gruppen/Zonen (bei Beibehaltung der oberen Reihenfolge), so dass die Anzahl der Artikel pro Gruppe/Zone in etwa gleich ist.

Das folgende idealisierte Beispiel in Abbildung 1 visualisiert die drei Zonen nach einer Bradfordizing-Analyse. Die Dokumentmenge umfasst in diesem Beispiel 450 Zeitschriftenartikel zu einem Forschungsthema, die in dem Fall auf insgesamt 39 unterschiedliche Zeitschriften verteilt sind (150 Artikel in jeder der drei Zonen). Das Kumulieren der Artikelzahlen nach dem Bradfordizing hat in diesem Fall ergeben, dass die ersten drei Zeitschriften zusammen 150 Artikel ergeben, also das erste Drittel der Gesamtdokumentenzahl von 450 Dokumenten. Diese ersten 150 Dokumente in den drei Zeitschriften definieren damit den Nukleus oder Core für dieses Topic. Die drei Zeitschriften werden „Core Journals“ oder Kernzeitschriften genannt. Für das zweite Drittel bzw. Zone 2 werden 9 (3*3 Zeitschriften) weitere Zeitschriften und für das dritte Drittel bzw. Zone 3 werden 27 (3*3*3 Zeitschriften) Zeitschriften benötigt, um die Menge von jeweils 150 Zeitschriftenartikeln zu erreichen.

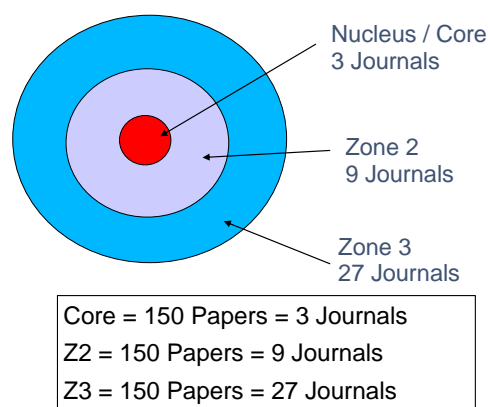


Abbildung 1: Ergebnis eines idealisierten Dokumentenpools für eine Fragestellung nach Bradfordizing. Einteilung in typische drei Bradford-Zonen (Core, Zone 2 und Zone 3). Die 450 Zeitschriftenartikel sind über 39 Zeitschriften verteilt.

Die Hauptidee der Anwendung von Bradfordizing in diesem Artikel ist es, ein beliebiges thematisch spezifiziertes Suchergebnis durch das einfach anwendbare Verfahren umzuorganisieren (Re-Ranking, vgl. Lin, 2008) und den Effekt dieses Re-Rankings zu evaluieren. Ziel ist es, nach dem Re-Ranking mehr relevante

Dokumente für ein Topic im ersten Drittel der Ergebnismenge zu listen als in den nachfolgenden Zonen. Dieser Zugewinn an relevanten Dokumenten im Kern soll als Bradfordizing-Effekt bezeichnet werden (siehe dazu Mayr, 2008).

Die Fokussierung auf Bradfordizing erscheint vielversprechend, weil dieses Verfahren per se universell und disziplinübergreifend angelegt ist und zuverlässig sowohl innerhalb einer Datenbank, einer Domäne (mehreren Datenbanken zu einem Fachgebiet) als auch zwischen Domänen beobachtet werden kann.

Bradfordizing liefert folgende unmittelbare Mehrwerte für den Nutzer:

- eine alternative Sicht auf Suchergebnisse, die nach Kernzeitschriften umorganisiert sind,
- eine alternative Sicht auf Publikationsquellen innerhalb eines Suchraums, die intuitiv näher am Forschungsprozess liegt als textstatistische Verfahren (z. B. best match) oder traditionelle boolesche Verfahren (exact match),
- eine vermutlich höhere fachliche Relevanz (topicality) der Dokumente nach dem Re-Ranking.

Aufgrund der Robustheit und Allgemeingültigkeit des BLS für die Verteilung von Forschungsliteratur ist davon auszugehen, dass das Bradfordizing gerade in föderierten und interdisziplinären Suchumgebungen mit unterschiedlichen Informationsbeständen praktisch operabel und gewinnbringend ist.

5 Forschungsfragen

Ziel der nachfolgenden Untersuchung ist es, eine umfangreiche Anzahl von standardisierten Fragestellungen in unterschiedlichen Datenbanken und Fachgebieten durch Bradfordizing zu analysieren. Dabei stellen sich folgende Fragen: Kann das nicht-textorientierte Rankingverfahren Bradfordizing ganz generell als ein Mehrwertdienst in Informationssystemen (z. B. Fachdatenbanken, Fachportalen mit mehreren Datenbanken) eingesetzt werden? Verbessert das Bradfordizing den Anteil der relevanten Dokumente im ersten Drittel (Core) einer Dokumentmenge gegenüber dem zweiten und dritten Drittel (Zone 2 bzw. Zone 3)? Zu diesem Zweck sollen der Re-Ranking-Ansatz Bradfordizing für Zeitschriftenartikel angewendet und in einem realen Suchszenario evaluiert werden. Des Weiteren wird geprüft, ob sich das Bradfordizing, das ursprünglich nur für Zeitschriftenliteratur beschrieben wurde, auch auf andere Dokumenttypen anwenden lässt.

Folgende Forschungsfragen lassen sich formulieren:

1. Welche Auswirkungen/Effekte hat die Anwendung von Bradfordizing als Re-Ranking-Ansatz für das Information Retrieval (IR)? Lassen sich die Effekte innerhalb eines klassischen IR-Tests evaluieren?
2. Lässt sich das Verfahren Bradfordizing als Re-Ranking-Ansatz auch auf andere Dokumenttypen und Szenarien übertragen? Konkret - kann Bradfordizing auch auf Monographien angewendet werden?

Im folgenden Abschnitt wird die Vorgehensweise der Evaluation erläutert. Eine detaillierte Beschreibung der angewendeten Methoden findet sich in Mayr (2010).

6 Design und Ablauf der Studie

In dieser Studie wurde Bradfordizing auf Dokumentenpools aus unterschiedlichen Fachdatenbanken aus den Bereichen Sozial- und Politikwissenschaften, Wirtschaftswissenschaften, Psychologie und Medizin zu verschiedenen Fragestellungen angewendet (siehe Tabelle 1). Für die Retrieval Tests wurden Daten aus zwei umfangreichen Evaluationsprojekten verwendet: zum einen Fragestellungen und Bewertungsdaten der jährlich

stattfindenden internationalen CLEF-Initiative (siehe <http://clef.iei.pi.cnr.it/>), zum anderen Daten, die im Rahmen des GESIS-Projekts KoMoHe im Jahr 2007 entstanden sind (siehe Mayr & Petras, 2008). Die beiden Evaluationsprojekte eignen sich besonders für diese Studie, da es sich bei den getesteten Dokumenten größtenteils um Zeitschriftenartikel und Monographien handelt, die bezüglich konkreter Fragestellungen intellektuell bewertet wurden. Für jede Fragestellung (im Kontext Information Retrieval Topic genannt) stehen jeweils ein Datensatz (Dokumentenpool) mit gefundenen Dokumenten und die entsprechenden binären Bewertungsdaten (relevant oder nichtrelevant) der Juroren zur weiteren Auswertung zur Verfügung (zum Verfahren siehe Harman & Voorhees, 2006). Die Topics bzw. die Dokumentenpools zu diesen Topics dienen jeweils als Grundlage für die Bradfordizing-Analyse (siehe dazu Tabelle 1).

	CLEF	KoMoHe
Zeitraum	2003-2007	2007
Anzahl der Topics	125	39
Domäne/Disziplin	Sozial- und Politikwissenschaften	Sozial- und Politikwissenschaften, Wirtschaftswissenschaften, Psychologie und Medizin
Bewertete Dokumente total	65.297	31.155
Zeitschriftenartikel bradfordized	18.112	17.432
Monographien bradfordized	11.045	4.900
Datenbanken involviert	1 (SOLIS)	6 (SOLIS, SoLit, USB Köln OPAC, Psynindex, World Affairs Online, Medline)

Tabelle 1: Übersicht über die analysierten Topics, Dokumente und Datenbanken in den IR-Experimenten (CLEF und KoMoHe)

Insgesamt wurden 164 intellektuell bewertete Dokumentenpools aus beiden Projekten analysiert. In diesen Dokumentenpools befanden sich mehr als 96.000 Dokumente, von denen mehr als 51.000 Dokumente durch das Verfahren Bradfordizing behandelt werden konnten.

In dieser Studie wird Bradfordizing auf die beiden Dokumenttypen Zeitschriftenartikel und Monographien angewendet. Da die Zeitschriftenartikel i. d. R. eine identifizierende Nummer der Zeitschrift (ISSN-Nummer) tragen, kann die anschließende Häufigkeitsanalyse (das eigentliche Bradfordizing) auf Basis der ISSN (International Standard Serial Number) erfolgen. Das gleiche Verfahren kann auf die ISBN-Nummer bei den Monographien angewendet werden. Die ISBN (International Standard Book Number) ist ein Identifier für Monographien und andere selbstständige Veröffentlichungen, in dem der Verlag, der die Publikation herausgibt, kodiert ist. Die Verlagsnummer ist eine ein- oder mehrstellige Zahl, die eindeutig einem Verlag zugeordnet ist. Dieser Verlags-Code wurde für die Häufigkeitsanalyse (Bradfordizing) verwendet.

Im Experiment wurden zunächst die beiden Dokumenttypen Zeitschriftenartikel und Monographien pro Topic aus den Dokumentenpools extrahiert und jeweils getrennt voneinander betrachtet. Abbildung 2 verdeutlicht den Prozess der Auswertung der evaluierten Topics aus den beiden Evaluationsprojekten KoMoHe und CLEF nach dem Verfahren Bradfordizing.

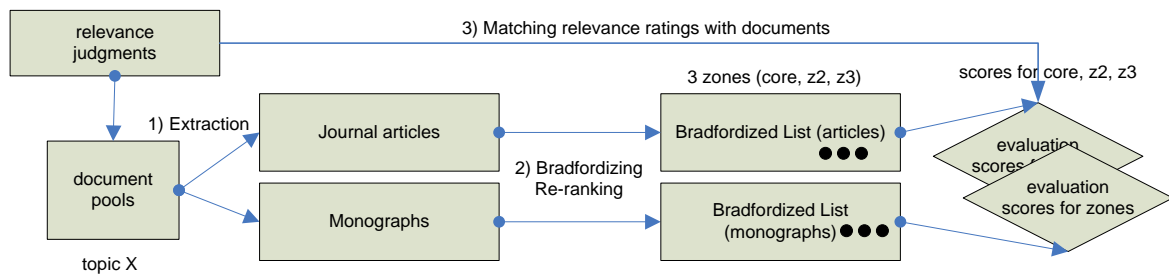


Abbildung 2: Auswertung der Relevanzbewertung im Re-Ranking-Experiment (KoMoHe und CLEF-Topics)

Die Analyse der Datensätze teilt sich in drei Schritte:

1. Die Dokumenttypen Zeitschriftenartikel (journal articles) und Monographien (monographs) werden aus dem ursprünglichen Datensatz extrahiert. Jedes Topic wird separat analysiert.
2. Anschließend wird die Dokumentmenge pro Topic nach dem Verfahren Bradfordizing (Häufigkeit einer Zeitschrift bzw. eines Verlages) umorganisiert und in drei etwa gleich große Zonen eingeteilt (core, z2 und z3).
3. Im nächsten Schritt wird die Verteilung der Relevanzbewertungen der Dokumente in diesen Zonen analysiert. Dazu werden die Dokumenten-IDs mit den Bewertungen zu den jeweiligen Topics abgeglichen und ausgewertet.

Im Anschluss an die Untersuchung liegen für jedes Topic und jeden Dokumenttyp die Anteile der relevanten bzw. nichtrelevanten Dokumente für die jeweilige Zone des Dokumentpools vor. Daraus lassen sich Precision-Werte berechnen. Die Precision misst das Verhältnis von gefundenen relevanten Dokumenten zu einer Suchanfrage an allen gefundenen Dokumenten aus einer Kollektion. Die Precision misst damit die Präzision bzw. Exaktheit eines Retrieval-Ergebnisses. Wenn beispielsweise nur relevante Dokumente aus der Kollektion zu einer Fragestellung gefunden werden, nimmt die Precision den Wert 1 an.

Es wurde pro Topic und Zone (core, z2 und z3) die durchschnittliche Precision berechnet. Zudem wurde die durchschnittliche Precision für das gesamte Dokumentset (baseline) berechnet (siehe Tabelle 2 als Beispiel).

	Retrieved	Relevant	Precision
Core	73	41	0.56 (P core)
Zone 2	65	25	0.38 (P z2)
Zone 3	70	14	0.20 (P z3)
Total	208	80	0.38 (P baseline)

Tabelle 2: Beispiel für die Berechnung der Precision-Werte für ein Topic

Folgende Hypothese lässt sich für die Auswertung der Precisionwerte in unserem Test formulieren:

Wenn der Anteil der relevanten Dokumente, gemessen in Precision (P), in allen drei Zonen gleich groß ist, dann hat das Verfahren Bradfordizing keinen Effekt auf die Verteilung der Relevanz im gesamten Dokumentpool. Wenn der Anteil der relevanten Dokumente in Zone Core niedriger ist als in Zone 2 oder Zone 3, dann verschlechtert sich das Suchergebnis. Wenn der Anteil der relevanten Dokumente im Core höher liegt als in den übrigen Zonen, was erwartet wird, dann verbessert das Verfahren das Suchergebnis und hat somit einen positiven Effekt auf die Suche.

7 Ergebnisse

Die Hauptergebnisse der Analyse der Dokumentverteilung nach Bradfordizing werden nachfolgend knapp am Beispiel der Zeitschriften zusammengefasst. Die Auswertungen zur Monographieliteratur finden sich in Mayr (2010).

- Die Analyse der Topics aus den Projekten CLEF und KoMoHe zeigt deutlich, dass Bradford-ähnliche Verteilungen für beide Datensätze (Datenbanken SOLIS, SoLit, USB Köln OPAC, Psynindex, World Affairs Online, Medline) und beide Dokumenttypen (Zeitschriftenartikel und Monographien) nachgewiesen werden können. Prinzipiell kann das Verfahren Bradfordizing aufgrund des Auftretens der Bradfordschen Gesetzmäßigkeit für beide Dokumenttypen auf Basis der rein quantitativen Analyse angewendet werden.

Der Einsatz von Bradfordizing soll das Suchergebnis komplett neu zusammensetzen und es möglichst signifikant verbessern. Ziel ist es, mehr relevante Dokumente für das Topic im ersten Drittel der Ergebnismenge zu listen. Im Mittelpunkt der Analyse steht der Vergleich der durchschnittlichen Precision-Werte der drei typischen Dokumentzonen (Bradford-Zonen) für bewertete Dokumentenpools. Die Auswertung der Relevanzurteile ergibt folgendes Ergebnis:

- Die Auswertung der Relevanzurteile für den Dokumenttyp Zeitschriftenartikel zu den 125 Topics der CLEF-Reihe zeigt ein deutliches Ergebnis. Die Dokumente, die nach Bradfordizing in die Kernzone (core) gerankt werden, sind im Mittel häufiger relevant für die Fragestellungen als Dokumente, die sich in Zone 2 oder Zone 3 befinden. Tabelle 3 zeigt die gemittelten Precision-Werte für 125 CLEF-Topics. Im Durchschnitt beträgt die Precision in der Kernzone 0,28. Die Precision sinkt in den folgenden Zonen deutlich ab (Zone 2 = 0,23 und Zone 3 = 0,19). Die baseline-Precision fasst die durchschnittliche Precision über alle Zeitschriftenartikel zusammen. Tabelle 4 zeigt die prozentuale Verbesserung der Precision der drei Zonen untereinander. Als Vergleichswert ist zusätzlich die baseline-Precision mitgeführt. Signifikante Verbesserungen sind jeweils durch das Sternsymbol gekennzeichnet (*). Es wird nicht zwischen unterschiedlichen Signifikanzniveaus unterschieden. Die Kernzone (core) enthält demnach über alle CLEF-Jahrgänge ca. 56 % mehr relevante Dokumente als Zone 3. Die Kernzone beinhaltet ebenfalls ca. 20 % mehr relevante Dokumente als die nachfolgende Zone 2. Zone 2 beinhaltet wiederum ca. 28 % mehr relevante Dokumente als Zone 3. Der Vergleich von Core und baseline-Precision fällt ebenfalls positiv für die Dokumente im Core aus. Die Zeitschriftenartikel im Core erzielen eine um knapp 22 % bessere Precision als der Durchschnitt über alle Zeitschriftenartikel.

Die Ergebnisse der CLEF-Evaluation können im KoMoHe-Datensatz bestätigt werden. Tabelle 5 zeigt die gemittelten Precision-Werte für 39 KoMoHe-Topics. Die 39 Topics zeigen die gleiche Tendenz: Die Dokumente der Kernzone sind signifikant häufiger relevant als Dokumente der Zone 2 und Zone 3. Im Durchschnitt beträgt die Precision in der Kernzone 0,40. Die Precision sinkt in den folgenden Zonen ab (Zone 2 = 0,37 und Zone 3 = 0,34). Die Kernzone (core) enthält demnach über alle KoMoHe-Tests ca. 17 % mehr relevante Dokumente als Zone 3 (siehe Tabelle 6). Die Kernzone beinhaltet ebenfalls ca. 9 % mehr relevante Dokumente als die nachfolgende Zone 2. Zone 2 beinhaltet wiederum ca. 7 % mehr relevante Dokumente als Zone 3. Der Vergleich von Core und baseline-Precision fällt ebenfalls positiv für die Dokumente im Core aus. Die Zeitschriftenartikel im Core erzielen eine um knapp 8 % bessere Precision als der Durchschnitt über alle Zeitschriftenartikel.

CLEF articles	Topics	P core	P Z2	P Z3	P baseline
2003	25	0.294	0.218	0.157	0.221
2004	25	0.226	0.185	0.134	0.179
2005	25	0.310	0.240	0.174	0.239
2006	25	0.288	0.267	0.244	0.265
2007	25	0.278	0.256	0.217	0.248

Tabelle 3: Durchschnittliche Precision für Zeitschriftenartikel nach Re-Ranking. Fünf CLEF-Jahrgänge (N=125 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline

CLEF articles	Core against Z3 in %	Core against Z2 in %	Z2 against Z3 in %	core against baseline in %
2003	86.56 (*)	34.57 (*)	38.63 (*)	32.65 (*)
2004	69.23 (*)	22.45	38.20	26.25 (*)
2005	78.03 (*)	29.05 (*)	37.95 (*)	29.52 (*)
2006	17.63	7.66	9.27	8.46
2007	28.18 (*)	8.31	18.35	11.77
Average 2003-2007	55.93 (*)	20.41 (*)	28.48 (*)	21.73 (*)

Tabelle 4: Durchschnittliche Precision-Verbesserung in % für Zeitschriftenartikel. Fünf CLEF-Jahrgänge (N=125 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline

KoMoHe Articles	Topics	P core	P Z2	P Z3	P baseline
Test1	15	0.292	0.261	0.245	0.265
Test2	12	0.215	0.202	0.192	0.202
Test3	12	0.700	0.644	0.587	0.642

Tabelle 5: Durchschnittliche Precision für Zeitschriftenartikel nach Re-Ranking. Drei KoMoHe-Tests (N=39 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline

KoMoHe articles	Core against Z3 in %	Core against Z2 in %	Z2 against Z3 in %	Core against baseline in %
Test1	18.82	11.75	6.32	9.84
Test2	11.58	6.16	5.11	6.12
Test3	19.32 (*)	8.67 (*)	9.80 (*)	9.00 (*)
Average Test1-3	16.57 (*)	8.86	7.08 (*)	8.32 (*)

Tabelle 6: Durchschnittliche Precision-Verbesserung in % für Zeitschriftenartikel. Drei KoMoHe-Tests (N=39 topics). Core, Zone 2 (Z2), Zone 3 (Z3) und baseline

Beim Vergleich dieser Ergebnisse zeigt sich, dass die Verbesserung der Precision zwischen den Zonen bei den KoMoHe-Tests nicht so deutlich ausfällt wie bei den CLEF-Topics (siehe Tabelle 4 und 6).

Die Dokumente der Kernzone zeigen für die meisten Testreihen signifikant höhere Precision als Dokumente der Zone 2 und Zone 3. Sowohl für die Zeitschriften als auch für die Monographien kann diese Aussage auf einer sehr breiten Basis von Themen und Fragestellungen an zwei unabhängigen Dokumentkorpora empirisch nachgewiesen werden. Das Verfahren Bradfordizing bzw. das Re-Ranking nach der Zeitschriftenhäufigkeit ist demnach für die zu einem Thema erfasste Literatur in Fachdatenbanken (insbesondere Zeitschriftenliteratur) aufgrund der signifikanten Precision-Verbesserung sehr vielversprechend.

8 Zusammenfassung

Die Besonderheit dieser Untersuchung liegt darin, dass für eine große Anzahl von sehr unterschiedlichen Fragestellungen die Konzentration der Literatur in vergleichsweise kleinen (unvollständigen) Dokumentkorpora untersucht wurde.

Die Precision für 125 getestete Topics in der Datenbank SOLIS erhöht sich nach Bradfordizing um durchschnittlich 56 % (Vergleich der Precision der Kernzeitschriften und der Zone 3-Zeitschriften). Diese signifikante Verbesserung durch Bradfordizing tritt auch für andere Datenbanken und Disziplinen ein. Die Precision für 39 getestete Topics in den Datenbanken SOLIS, SoLit, USB Köln OPAC, Psyn dex, Econis, World Affairs Online, Medline erhöht sich nach Bradfordizing um durchschnittlich 17 % (Vergleich der Precision der Kernzeitschriften und der Zone 3-Zeitschriften).

Die Dokumente der Kernzone zeigen für die meisten Testreihen signifikant höhere Precision als Dokumente der Zone 2 und Zone 3. Sowohl für die Zeitschriften als auch für die Monographien kann diese Aussage auf einer sehr breiten Basis von Themen und Fragestellungen an zwei unabhängigen Dokumentkorpora empirisch nachgewiesen werden. Das Verfahren Bradfordizing bzw. das Re-Ranking nach der Zeitschriftenhäufigkeit ist demnach für die erfasste Literatur in Fachdatenbanken (insbesondere Zeitschriftenliteratur) zu einem Thema aufgrund der signifikanten Precision-Verbesserung sehr vielversprechend. Gestützt werden die empirischen Ergebnisse durch eine qualitative Fallstudie die ebenfalls im Rahmen der Dissertation durchgeführt wurden (siehe dazu Mayr, 2009).

Zusätzlich zu den Zeitschriften wurde ein Teil der CLEF- und KoMoHe-Topics auch für die Monographienliteratur untersucht (siehe detaillierter in Mayr, 2010). Die bei Worthen (1975) beschriebene Übertragbarkeit der Bradfordschen Gesetzmäßigkeit auf die Monographienliteratur kann durch diese Studie für die analysierten Topics auf einer sehr breiten Basis bestätigt werden. Es zeigt sich, dass sich das in dieser Arbeit entwickelte Vorgehen der Analyse der Verlags-Codes aus der ISBN der Monographie-Datensätze als tragbares Verfahren erweist.

Sehr abstrakt ausgedrückt soll Bradfordizing durch die Identifikation von Kernbereichen als ein Kompensationsverfahren für erweiterte Suchräume (mehrere Datenbanken in einer Suchanfrage) eingesetzt werden. Der Heterogenität und der Vielzahl der Datenbanken wird als Ausgleich ein quasi vereinheitlichendes Prinzip entgegengesetzt, das Kernbereiche identifiziert, die über Datenbanken und Domänen hinweg existieren.

Folgende beiden Thesen können als Ergebnisthesen formuliert werden:

1. Die Anwendung des Bradfordizings bzw. das Re-Ranking nach Kernzeitschriften für thematische Dokumentmengen führt zu signifikanten Verbesserungen der Precision zwischen den drei Zonen (Core, Zone 2 und Zone 3). Die Kernzeitschriften (Core) beinhalten signifikant mehr relevante Dokumente als

Zeitschriften der Zone 2 oder der Zone 3. Der größte Precision-Gewinn ergibt sich zwischen Core und Zone 3-Zeitschriften.

2. Bradfordizing für thematisch konzentrierte Dokumentmengen lässt sich erfolgreich auf Monographien (Verlag als Selektionskriterium) übertragen. Die Anwendung des Bradfordizings bzw. das Re-Ranking nach Verlagen für Monographien führt zu geringeren Verbesserungen der Precision zwischen den drei Zonen (Core, Zone 2 und Zone 3). Der größte Precision-Gewinn ergibt sich auch hier wie bei den Zeitschriften zwischen Core und Zone 3-Verlagen.

9 Ausblick

Die Ergebnisse dieser Arbeit gehen in das GESIS-Projekt „Retrieval-Mehrwertdienste zur Weiterentwicklung wissenschaftlicher Fachportale wie vascoda und sowiport. Suchexpandierung und Re-Ranking“ (kurz IR-Mehrwertdienste) ein, das im Januar 2009 gestartet wurde. Das Projekt IR-Mehrwertdienste² (DFG-Förderung INST 658/6-1) fokussiert auf die Implementation und Evaluation von drei Mehrwertdiensten (u. a. Bradfordizing), die die Informationssuche in Digitalen Bibliotheken verbessern sollen (siehe dazu Mayr et al., 2011). Abbildung 3 zeigt ein Beispiel eines nach Bradfordizing gerankten Suchergebnisses im Prototypensystem des Projekts IRM³.

Information Retrieval Value-added Services

Query:

Only show metadata sets which include an abstract ☒

Applied filters: abstract: [* TO *]

Suggest search terms	Rerank the result list
Controlled vocabulary Thesaurus Sozialwissenschaften <input type="button" value="v"/>	Rerank method Bradfordizing <input type="button" value="v"/>
Automatic query expansion <input type="checkbox"/>	

Total hits: 373

1. [Frankreichs solidarische securite sociale in der Krise](#) (1997)
journalarticle in [Sozialer Fortschritt unabhängige Zeitschrift für Sozialpolitik](#) Jg. 46 H. 1-2 (0038-609X) by Kaufmann, Otto
2. [Das belgische Sozialversicherungssystem auf dem Prüfstand : die Antworten auf die aktuellen Herausforderungen](#) (1997)
journalarticle in [Sozialer Fortschritt unabhängige Zeitschrift für Sozialpolitik](#) Jg. 46 H. 1-2 (0038-609X) by Jorens, Yves
3. [Der italienische Sozialstaat zwischen Krise und Reform](#) (1997)
journalarticle in [Sozialer Fortschritt unabhängige Zeitschrift für Sozialpolitik](#) Jg. 46 H. 1-2 (0038-609X) by Hohnerlein, Eva-Maria

Interactive query enhancement

[term cloud](#) ▼

[search term suggestions](#) ▼

[central journals](#) ▼

- [Sozialer Fortschritt](#) (31)
- [WSI Mitteilungen](#) (21)
- [Aus Politik und Zeitgeschichte](#) (20)
- [Deutsche Rentenversicherung](#) (18)
- [Blätter für deutsche und internationale Politik](#) (13)
- [Zeitschrift für Sozialreform](#) (12)
- [Arbeit und Beruf](#) (10)
- [Deutscher Bundestag. Ausschuß für Arbeit, Kurzprotokoll](#) (8)
- [Soziale Sicherheit](#) (8)
- [Mitteilungen aus der Arbeitsmarkt- und Berufsforschung](#) (8)
- [Utopie kreativ](#) (7)
- [Neue Praxis](#) (7)
- [Theorie und Praxis der sozialen Arbeit](#) (7)

Abbildung 3: Beispiel eines neugerankten Ergebnisses (nach Bradfordizing) im IRM-Prototyp

² <http://www.gesis.org/irm>

³ <http://www.gesis.org/beta/prototypen/irm/>

Literatur

1. Bates, Marcia J. (1989): The Design of Browsing and Berrypicking Techniques for the Online Search Interface. In: Online Review 13, Nr. 5, S. 407-424. URL: <http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>
2. Bates, Marcia J. (2002): Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution. S. 137-150. In: Bruce, Harry; Fidel, Raya; Ingwersen, Peter; Vakkari, Pertti (Hrsg.): Fourth International Conference on Conceptions of Library and Information Science (CoLIS 4) URL: http://www.gseis.ucla.edu/faculty/bates/articles/Searching_Bradford-m020430.html
3. Bradford, Samuel C. (1934): Sources of information on specific subjects. In: Engineering 137, Nr. 3550, S. 85-86
4. Bradford, Samuel C. (1948): Documentation. London: Lockwood. 156 S.
5. Egghe, Leo; Rousseau, Ronald (1990): Introduction to Informetrics: quantitative methods in library, documentation and information science: Elsevier Science Publishers. URL: <http://eprints.rclis.org/archive/00003297/>
6. Garfield, Eugene (1990): How ISI Selects Journals for Coverage: Quantitative and Qualitative Considerations. In: Current Contents 13, Nr. 22, S. 5-13. URL: <http://www.garfield.library.upenn.edu/essays/v13p185y1990.pdf>
7. Garfield, Eugene (1996): The Significant Scientific Literature Appears In A Small Core Of Journals. In: The Scientist 10, Nr. 17, S. 13. URL: [http://www.garfield.library.upenn.edu/commentaries/tsv10\(17\)p13y090296.html](http://www.garfield.library.upenn.edu/commentaries/tsv10(17)p13y090296.html)
8. Harman, Donna K.; Voorhees, Ellen M. (2006): TREC: An Overview. In: Annual Review of Information Science and Technology 40, S. 113-156
9. Hood, William W.; Wilson, Concepción S. (2001): The scatter of documents over databases in different subject domains: how many databases are needed? In: Journal of the American Society for Information Science and Technology 52, Nr. 14, S. 1242-1254
10. Krause, Jürgen (2004): Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie 51, Nr. 2, S. 76 - 89
11. Krause, Jürgen (2008): Total Package Design für Digitale Bibliotheken und Fachinformation. S. 185-206. In: Hutzler, Evelinde; Schröder, Albert; Schweikl, Gabriele (Hrsg.): Bibliotheken gestalten Zukunft. Kooperative Wege zur Digitalen Bibliothek. Dr. Friedrich Geißelmann zum 65. Geburtstag. Göttingen: Universitätsverlag Göttingen. URL: <http://epub.uni-regensburg.de/4564/>
12. Krause, Jürgen; Niggemann, Elisabeth; Schwänzl, Roland (2003): Normierung und Standardisierung in sich verändernden Kontexten: Beispiel Virtuelle Fachbibliotheken. In: ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie 50, Nr. 1, S. 19 - 28
13. Lin, Jimmy (2008): PageRank without hyperlinks: Reranking with PubMed related article networks for biomedical text retrieval. In: BMC Bioinformatics 9. URL: <http://www.biomedcentral.com/1471-2105/9/270>
14. Mayr, Philipp (2008): An evaluation of Bradfordizing effects. In: Kretschmer, H.; Havemann, F. (Hrsg.): Proceedings of WIS 2008, Berlin, Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting. Humboldt-Universität zu Berlin URL: <http://www.collnet.de/Berlin-2008/MayrWIS2008ebe.pdf>
15. Mayr, Philipp (2010): Information Retrieval-Mehrwertdienste für Digitale Bibliotheken: Crosskonkordanzen und Bradfordizing. GESIS-Schriftenreihe 5.
16. Mayr, Philipp; Mutschke, Peter; Schaer, Philipp; Sure, York (2011 to appear): Science Models as Value-Added Services for Scholarly Information Systems. In: Scientometrics
17. Mayr, Philipp; Mutschke, Peter; Petras, Vivien (2008): Reducing semantic complexity in distributed digital libraries: Treatment of term vagueness and document re-ranking. In: Library Review 57, Nr. 3, S. 213-224. URL: http://www.ib.hu-berlin.de/~mayr/arbeiten/mayr-et-al_LR08.pdf

18. Mayr, Philipp; Petras, Vivien (2008): Cross-concordances: terminology mapping and its effectiveness for information retrieval. In: 74th IFLA World Library and Information Congress. Québec, Canada URL: http://www.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf
19. Mutschke, Peter (2004): Autorennetzwerke: Netzwerkanalyse als Mehrwertdienst für Informationssysteme. S. 141-162. In: Bekavac, Bernard; Herget, Josef; Rittberger, Marc (Hrsg.): Information zwischen Kultur und Marktwirtschaft. Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004). Chur: UVK Verlagsgesellschaft mbH. (Schriften zur Informationswissenschaft, Band 42) URL: <http://www.inf-wiss.uni-konstanz.de/infwiss/download/cc-isi04-art9.pdf>
20. Nacke, Otto (1979): Informetrie: ein neuer Name für eine neue Disziplin, Nachrichten für Dokumentation 30, Nr. 6, S. 219-226.
21. Newman, M. E. J. (2005): Power laws, Pareto distributions and Zipf's law. In: Contemporary Physics 46, Nr. 5, S. 323-351. URL: <http://arxiv.org/abs/cond-mat/0412004>
22. Umstätter, Walther (2005): Anmerkungen zu Birger Hjørland und Jeppe Nicolaisen: Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". In: LIBREAS, Nr. 3. URL: http://www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe3/008ums.htm
23. Vickery, Brian C. (1948): Bradford's law of scattering. In: Journal of Documentation 4, Nr. 3, S. 198-203
24. Wagner-Döbler, Roland (1997): Time dependencies of Bradford distributions: structures of journal output in Twentieth century logic and Nineteenth century mathematics. In: Scientometrics 39, Nr. 3, S. 231-252
25. White, Howard D. (1981): 'Bradfordizing' search output: how it would help online users. In: Online Review 5, Nr. 1, S. 47-54
26. Wilkinson, E. A. (1972): The ambiguity of Bradford's law. In: Journal of Documentation 28, Nr. 2, S. 122-130
27. Worthen, D. B. (1975): The application of Bradford's law to monographs. In: Journal of Documentation 31, Nr. 1, S. 19-25